BEHAVIORAL IMPEDIMENTS TO VALUING ANNUITIES: COMPLEXITY AND CHOICE BRACKETING

Jeffrey R. Brown, Arie Kapteyn, Erzo F. P. Luttmer, Olivia S. Mitchell, and Anya Samek*

Abstract—This paper examines two behavioral factors that diminish people’s ability to value a lifetime income stream or annuity, drawing on a randomized experiment with about 4,000 adults in a U.S. nationally representative sample. We find that increasing the complexity of the annuity choice reduces respondents’ ability to value the annuity, measured by the difference between the sell and buy values they assign to the annuity. When we limit narrow choice bracketing by inducing people to think first about how quickly or slowly to spend down assets in retirement, their ability to value an annuity increases.

I. Introduction

ANNUITIES can be a valuable form of insurance against the possibility of exhausting financial resources or having to severely curtail retirement consumption. Nevertheless, there is relatively little demand for these insurance products (Mitchell, Piggott, & Takayama, 2011; Poterba, Venti, & Wise, 2011). A voluminous literature reviewed in Brown (2009) explores rational explanations for why observed levels of annuitization are much lower than predicted by standard optimizing models such as those by Yaari (1965) and Davidoff, Brown, and Diamond (2005). Recent contributions to this literature include several papers that combine multiple deviations from the standard optimizing framework. For instance, Ameriks et al. (2011, 2020) and Lockwood (2012, 2018) explain observed low annuity demand using structural models that combine a precautionary savings motive (for long-term-care expenses when there is public care aversion) with a bequest motive. Reichling and Smetter (2015) do so as well by introducing stochastic mortality and correlated uninsured health care costs. Peltenburg, Nijman, and Werker (2017) show that medical expenditure risk can rationalize low observed annuitization levels early in retirement, but not why many older people fail to buy annuities. Finally, Laitner, Silverman, and Stolyarov (2018) show analytically how the presence of implicit longevity insurance provided by Medicaid nursing home care can crowd out demand for annuities for the lower and middle classes.

A different strand of literature explores whether behavioral factors help explain low observed levels of annuitization. Several hypothetical choice experiments suggest that behavioral factors influence the demand for annuities, including studies showing that framing of the annuity choice affects the demand for annuities (Brown et al., 2008, 2013; Beshears et al., 2014; Brown, Kapteyn, & Mitchell, 2016; Merkle, Schreiber, & Weber, 2017; and Bockweg et al., 2018). Similar findings emerge in incentivized laboratory settings (Agniew et al., 2008; Gazzale & Walker, 2011). Another source of evidence is research demonstrating that individuals in a hypothetical choice setting provide widely divergent valuations for small increases versus small decreases in annuitization amounts (Brown et al., 2017). This latter result is consistent with people having trouble assessing the value of an annuity stream and therefore requiring a high selling price and offering a low buying price, as they are reluctant to trade what they do not understand. There is also suggestive evidence from nonhypothetical choices that points to behavioral mechanisms. For instance, in ten Swiss firms, Bütler and Teppa (2007) show that annuitization rates are much higher on average in the firms that offer an annuity as the default payout option than in the one firm paying out a lump sum as the default. This finding suggests that annuitization rates are influenced by the default, implying a deviation from a standard rational model. Similarly, Hagen, Hallberg, and Lindquist (2018) show that a nudge affects annuitization decisions of Swedish pensioners. Other papers finding patterns in observed annuitization choices suggestive of deviations from rational choice models include Hurd and Panis (2006), Chalmers and Reuter (2012), Previtero (2014), and Fitzpatrick (2015). Shepard (2011) and Bronshtein et al. (2016) use arbitrage arguments to show that for many people, the annuitization decision implicit in when to claim Social Security benefits cannot be fully explained by a standard rational model.

Although rational models can be constructed to match the low observed demand for annuities, our take from the literature on the annuity puzzle is that behavioral factors remain operative. In short, we share Brown’s (2009, 185) assessment that while “it is possible to generate more limited annuitization by extending the rational model in several directions, such an approach does not seem to provide the complete answer to the puzzle” of low observed levels of annuitization. Similarly, Benartzi, Previtero, and Thaler (2011, 61) conclude
that the “tiny market share of individual annuities should not be viewed as an indicator of underlying preferences but rather as a consequence of institutional factors about the availability and framing of annuity options.”

Many studies find that behavioral factors influence annuitization decisions, yet relatively little is known about the mechanisms driving this behavior. Brown et al. (2008, 2013) report that presenting annuities in terms of the consumption streams they generate leads to higher annuity demand than presenting annuities as investment products. Brown et al. (2008) suggest that the adoption of a narrow decision frame, also referred to as choice bracketing (Thaler, 1985; Read, Loewenstein, & Rabin, 1999), may drive this finding: that is, people evaluate annuities based on the return and variance of the payouts in isolation rather than by focusing on the level and variance of the consumption stream flowing from the annuity (which is what matters for utility). It remains a leap of faith, however, to infer that the choice is more rational simply because demand is higher. Brown et al. (2017) establish that the deviation from rational choice, measured by the gap between peoples’ sell-versus-buy prices for annuities, is lower for individuals with better cognition scores. The authors take this as suggestive evidence that valuing annuities is cognitively challenging because it is a complex task. Nevertheless, they do not claim that this is causal evidence of a mechanism, as they lack exogenous variation in the complexity of the annuitization decision.

In this paper, we produce stronger evidence on behavioral mechanisms that may affect the annuitization decision. Rather than asking for a respondent’s own hypothetical annuitization decision, we first describe a vignette in which a hypothetical person faces a choice between a $100 per month decrease in Social Security benefits versus receiving a lump-sum amount. We ask each respondent what the vignette person should choose and repeat the question for various values of the lump sums until we find the lump sum deemed equivalent in value to a $100 per month increase in the Social Security annuity. We call this lump-sum amount the “sell” valuation because the respondent advises the vignette person to sell a $100-a-month annuity for this lump sum. At a different point in the experiment, we ask each respondent to advise the same vignette person on a choice between a $100 per month decrease in Social Security benefits versus paying a lump sum. The lump-sum amount that is valued as much as the decrease in benefits is the “buy” valuation, as it represents the amount of money the respondent advises the vignette person to pay to avoid forfeiting a $100-per-month annuity. We refer to the absolute difference between the log sell valuation and the log buy valuation as the sell-buy spread, and we use this to measure deviations from rational decision-making.

We introduce two experimental interventions to test for two types of behavioral impediments to valuing annuities. First, we vary the ease by which an annuity stream can be valued, which we refer to as the complexity of the annuitization choice. Valuing an annuity stream is more difficult when there is greater uncertainty about longevity. We experimentally manipulate this uncertainty by telling the respondent what longevity information the vignette person received from a doctor. Valuing an annuity is also more difficult when the description of the annuity contains additional information that turns out to be irrelevant but nevertheless requires effort to process. This is an alternative means by which we vary complexity. Second, we independently randomize whether the respondent receives information about the benefits and drawbacks of spending down nonannuitized wealth during retirement more rapidly versus more slowly. This intervention occurs before the respondent advises the vignette person about annuitization. The purpose of the intervention is to induce people to think about the consumption consequences of holding an annuity during retirement. The “consequence message” intervention therefore has the potential to be a new instrument (besides framing) to reduce the narrow choice bracketing that Brown et al. (2008) identified as a behavioral mechanism.

Our experiment yields two main findings. First, we show that greater complexity causes the sell-buy spread to increase, indicating that complexity associated with annuities reduces people’s ability to assess the value of an annuity. This is the

---

1 Policy risk reduces people’s valuation of the stream of Social Security benefits (Luttmer & Samwick, 2018), which should reduce both the buy and sell valuation, leaving their difference unaffected.

2 As described below, we have included additional experimental interventions to test for anchoring and to test whether results are robust. All of these experimental interventions are orthogonal to the two main interventions designed to test for behavioral impediments to valuing annuities.

3 Our reference to complexity differs from a common use of the term when describing smaller or larger choice sets (e.g., Carvalho & Silverman, 2019).
Evidence that complexity impedes annuitization decisions has the important implication that reducing complexity can improve individuals’ annuitization decisions. While it may be possible to make the decision less complex by presenting information about the annuity more clearly, we stress that much of the complexity is inherent in the annuitization decision itself: people need to jointly evaluate how much they will consume each future year with and without the annuity, how much they care about consumption fluctuations, and the probability that they will be alive in each future year. No matter how well the decision is presented, it remains a complex task. We do find that inducing people to consider the consequences of annuitization decisions for their consumption streams enables them to better assess the value of an annuity. This is important because it provides clear guidance on how annuitization decisions should be presented. Still, while the consequence message limits the degree to which choice bracketing acts as an impediment to valuing an annuity, we emphasize that the sell-buy spread remains substantial even for those exposed to the consequences message.

The rest of the paper proceeds as follows. Section II describes our methodology and explains our experimental design. In section III, we present our empirical findings, and section IV concludes.

II. Methodology and Experimental Design

A. Understanding America Study

Our experiment uses the UAS, a probability-based Internet panel of about 6,000 adults (age 18 and over) representative of the U.S. population. Panel members are recruited exclusively through address-based sampling, in which invitation letters are sent to randomly selected households using address lists obtained from the U.S. Postal Service. This provides a broadly representative sample, since individuals lacking prior access to the Internet are provided with a tablet and broadband Internet. In addition, the UAS contains small oversamples (about 5% each) of Native Americans and residents of Los Angeles County. Our experimental module was fielded between June and October 2016, and all UAS panel members at the time were invited to participate. Panel members received $10 for completing the survey, which took an average of fourteen minutes, and they could also receive additional earnings depending on their answers to quiz questions. Of the 5,521 invited panel members, 83.2% opened the link to the survey. Of those who opened the link, 99.1% completed exclusively as an annuity is particularly valuable to people who would otherwise underannuitize.

Evidence that complexity impedes annuitization decisions has the important implication that reducing complexity can improve individuals’ annuitization decisions. While it may be possible to make the decision less complex by presenting information about the annuity more clearly, we stress that much of the complexity is inherent in the annuitization decision itself: people need to jointly evaluate how much they will consume each future year with and without the annuity, how much they care about consumption fluctuations, and the probability that they will be alive in each future year. No matter how well the decision is presented, it remains a complex task. We do find that inducing people to consider the consequences of annuitization decisions for their consumption streams enables them to better assess the value of an annuity. This is important because it provides clear guidance on how annuitization decisions should be presented. Still, while the consequence message limits the degree to which choice bracketing acts as an impediment to valuing an annuity, we emphasize that the sell-buy spread remains substantial even for those exposed to the consequences message.

The rest of the paper proceeds as follows. Section II describes our methodology and explains our experimental design. In section III, we present our empirical findings, and section IV concludes.

II. Methodology and Experimental Design

A. Understanding America Study

Our experiment uses the UAS, a probability-based Internet panel of about 6,000 adults (age 18 and over) representative of the U.S. population. Panel members are recruited exclusively through address-based sampling, in which invitation letters are sent to randomly selected households using address lists obtained from the U.S. Postal Service. This provides a broadly representative sample, since individuals lacking prior access to the Internet are provided with a tablet and broadband Internet. In addition, the UAS contains small oversamples (about 5% each) of Native Americans and residents of Los Angeles County. Our experimental module was fielded between June and October 2016, and all UAS panel members at the time were invited to participate. Panel members received $10 for completing the survey, which took an average of fourteen minutes, and they could also receive additional earnings depending on their answers to quiz questions. Of the 5,521 invited panel members, 83.2% opened the link to the survey. Of those who opened the link, 99.1% completed

Evidence that complexity impedes annuitization decisions has the important implication that reducing complexity can improve individuals’ annuitization decisions. While it may be possible to make the decision less complex by presenting information about the annuity more clearly, we stress that much of the complexity is inherent in the annuitization decision itself: people need to jointly evaluate how much they will consume each future year with and without the annuity, how much they care about consumption fluctuations, and the probability that they will be alive in each future year. No matter how well the decision is presented, it remains a complex task. We do find that inducing people to consider the consequences of annuitization decisions for their consumption streams enables them to better assess the value of an annuity. This is important because it provides clear guidance on how annuitization decisions should be presented. Still, while the consequence message limits the degree to which choice bracketing acts as an impediment to valuing an annuity, we emphasize that the sell-buy spread remains substantial even for those exposed to the consequences message.

The rest of the paper proceeds as follows. Section II describes our methodology and explains our experimental design. In section III, we present our empirical findings, and section IV concludes.

II. Methodology and Experimental Design

A. Understanding America Study

Our experiment uses the UAS, a probability-based Internet panel of about 6,000 adults (age 18 and over) representative of the U.S. population. Panel members are recruited exclusively through address-based sampling, in which invitation letters are sent to randomly selected households using address lists obtained from the U.S. Postal Service. This provides a broadly representative sample, since individuals lacking prior access to the Internet are provided with a tablet and broadband Internet. In addition, the UAS contains small oversamples (about 5% each) of Native Americans and residents of Los Angeles County. Our experimental module was fielded between June and October 2016, and all UAS panel members at the time were invited to participate. Panel members received $10 for completing the survey, which took an average of fourteen minutes, and they could also receive additional earnings depending on their answers to quiz questions. Of the 5,521 invited panel members, 83.2% opened the link to the survey. Of those who opened the link, 99.1% completed
both annuity valuation questions, for an overall response rate of 82.4% (4,549 respondents).

The UAS gathers information on demographic characteristics for all respondents, as well as detailed measures of cognitive capabilities and financial literacy (the latter for about 90% of respondents). Given that cognitive ability and financial literacy are important predictors of responses to annuity questions, we limit our analysis sample to observations with nonmissing measures of cognitive ability and financial literacy. In addition, we exclude 0.5% of observations with missing values for any demographic characteristics. The final analysis sample has 4,060 observations (89.2% of the total number of respondents who completed both the annuity sell and buy questions).

We recognize that a drawback of hypothetical choice data is that people may not put as much effort into making decisions as they might in real-life situations. As a result, their answers may contain more measurement error than would be true in the real world. Nevertheless, it seems unlikely that people can fully overcome cognitive biases simply by exerting more effort. Moreover, concerns about the reliability of willingness-to-pay responses in the UAS are allayed by Mas and Pallais (2017), who show that the distribution of willingness-to-pay responses in the UAS closely matches the willingness-to-accept responses obtained in the UAS closely matches the willingness-to-pay distribution from a similar field experiment. In our case, using hypothetical choice data has the important advantage that we can elicit both a willingness to pay and a willingness to accept for the same person, permitting us to measure deviations from rational decision making. We know of no field setting that allows for the simultaneous measurements of willingness to pay and a willingness to accept for an annuity for the same person. Moreover, in our settings, we observe the valuations of all respondents, in contrast to most revealed preference approaches where only the valuations of marginal individuals can be observed and the valuations of inframarginal persons can only be bounded, absent functional form assumptions.

Online appendix table A1 provides summary statistics for our baseline sample and compares it to the Current Population Survey (CPS) of the same year. Compared to the CPS, our sample overrepresents respondents between the ages of 35 and 65 by 11 percentage points, females by 6 percentage points, married respondents by 7 percentage points, non-Hispanic whites by 11 percentage points, individuals with more than a high school education by 16 percentage points, households with annual incomes above $75,000 by 3 percentage points, households with two or fewer members by 10 percentage points, and households with no children by 5 percentage points. While these differences are generally statistically significant, the two samples are reasonably similar in terms of economic magnitudes, with the absolute difference in the fraction of respondents in a category being 5 percentage points on average across the 25 demographic categories listed in the table. As such, we consider our sample to be broadly representative of the U.S. adult population.

B. Experimental Context

Rather than describing an unfamiliar hypothetical annuity product, we use Social Security benefits as the context for the analysis of payout annuities. Specifically, we asked respondents to make trade-offs between receiving higher or lower Social Security benefits (a change in a real annuity stream) and paying or receiving different one-time payments (lump sums). Our setting is policy relevant because past discussions of pension reforms around the world, including in the United States, have included proposals to offer workers lump-sum payments in exchange for a reduction in their annuitized pension benefits (Maurer et al., 2018). Several U.S. corporations have also recently offered to buy back defined benefit pension annuities from retirees in exchange for lump sums (Wayland, 2012).

C. Elicitation of the Valuation of an Annuity Stream

Throughout the experiment, we use vignettes to describe trade-offs and ask respondents to give the hypothetical vignette person advice about annuitization decisions. This approach has several attractive features. First, we can directly manipulate the complexity of the annuitization decision by using different experimental treatments. Second, we control for the respondent’s own characteristics: unlike making a decision for one’s own situation (as in Brown et al., 2017), we need not worry about factors such as liquidity constraints or private knowledge that the respondent may have about his or her own situation.

The vignette person in the control condition is described as follows:

Mr. Jones is a single, 60-year old man with no children. He will retire and claim his Social Security benefits at 65. When he retires, he will have $100,000 saved for his retirement, and he will receive $[SSB] in monthly Social Security benefits. Based on his current health and family history, doctors have told Mr. Jones that he will almost certainly be alive at age 75 but almost certainly will not live beyond age 85.

The gender and name of the vignette person are experimentally varied among respondents. The variable $[SSB]$ represents the vignette person’s monthly Social Security benefits, randomized with equal probability across respondents to $800, $1,200, $1,600, and $2,000.

Our main outcome of interest is the respondent’s advice for how the hypothetical vignette person should trade off annuitized wealth and lump-sum amounts at retirement. All respondents answer a series of questions that elicit either the

show you some examples of persons who have to make a decision about money and we will ask you to help them make the decision.”
equivalent variation (EV) of a $100 increase in monthly Social Security benefits or the EV of a $100 decrease in monthly Social Security benefits. Each respondent is asked both questions, and the order in which they are asked is randomized. The valuation of a $100 increment in the annuity stream is elicited by asking a series of questions of the form:

What should Mr. Jones do?

1. Receive a Social Security benefit of $[SSB + 100] per month starting at age 65.
2. Receive his expected Social Security benefit of $[SSB] per month and receive a one-time payment of $[LS]$ from Social Security at age 65.

The $100 increment in benefits of $[SSB + 100]$ is displayed as a single number on the screen. The variable LS represents the lump-sum amount that is traded off, randomized across respondents to start at $10,000, $20,000, or $30,000. The question is subsequently asked four more times for different values of LS. For example, if the person declines a $20,000 lump sum, we infer that the valuation must exceed $20,000, so for the next question, we use a higher value of LS: $60,000. If the person accepts the $20,000 lump sum, we would use a lower value of LS. Next, if the person accepts the $60,000 lump sum, we infer that the valuation must lie below $60,000, and we ask the question three more times to further reduce the difference between the lower and upper bound of the person’s valuation of the $100 increment in the annuity stream. The exact sequence of values for LS is shown in the survey instrument in the online appendix. We refer to this question as the “sell” version, because the person receives a payment in exchange for a smaller annuity stream.

The valuation of a $100 decrease in the annuity stream is elicited by asking a series of questions of the form:

What should Mr. Jones do?

2. Receive his expected Social Security benefit of $[SSB] per month and make a one-time payment of $[LS]$ to Social Security at age 65.

As before, the question is asked five times for different values of LS until we can place the respondent’s valuation of the annuity into one of 32 bins. We refer to this question as the buy version, because the person is making a payment in exchange for a larger annuity stream.

Given that a $100 change in the annuity stream is small relative to the average monthly benefit of $1,400, a rational respondent should value this change approximately the same, whether it is an increase or a decrease. We therefore take the absolute difference of the sell and buy valuations to measure the deviation from rational decision making.

D. Experimental Design

Our experiment consists of a $3 \times 2$ between-subjects design, summarized in table 1. First, we experimentally vary what we refer to as the complexity of the vignette in one of two ways: increasing the uncertainty associated with length of life (Complexity: Wide age range treatment), or adding extraneous information to the vignette that is not relevant to the decision (Complexity: Added information treatment). For example, control group respondents are told that the vignette person will “almost certainly be alive at age 75 but almost certainly will not live beyond age 85.” By contrast, respondents in the Complexity: Wide age range treatment are told that the vignette person “has an 80% chance of being alive at age 70, a 50% chance of being alive at age 80, a 20% chance of being alive at age 90, and a 10% chance of being alive at age 95.” Determining the value of an annuity is a more complex task when the variation in possible ages of death is more dispersed, as is the case in this second vignette. The extraneous information added to the Complexity: Added information treatment includes information about Social Security qualification rules and describes why the vignette person qualified. Here the increased complexity requires the respondent to think about the additional information and determine whether it is relevant.

Second, prior to the advice decision, in half of the treatments, we additionally provide a message about the...
consequences of spending down retirement savings (Consequence message). This message describes an interaction between a different vignette person and his or her financial adviser. In this interaction, the adviser describes the benefits and drawbacks of spending down savings relatively quickly (more likely to be able to use money in one’s lifetime but running a larger risk of running out of money while alive) versus relatively slowly (less likely to run out of money but running a larger risk of not getting to enjoy one’s money in one’s lifetime). This message is framed as neutrally as possible and designed to encourage the respondent to avoid narrow choice bracketing: by inducing respondents to think about the problem of how to spend down wealth in retirement, we intended that respondents consider the annuitization decision and the asset decumulation decisions jointly rather than as disjoint decisions. To ensure that respondents pay attention to the message, respondents are further told that at the end of the message, they will be asked two questions about the facts in the story and will receive an additional $1 for each question they answer correctly. These factual questions are two multiple-choice questions about the financial adviser’s explanation of the benefits and drawbacks under each scenario (spending down slowly or quickly). Of the respondents who are posed these two factual questions, 63% answer both correctly, 27% answer one correctly, and 10% answer neither correctly.

In summary, all respondents are asked to give advice to a primary vignette person about buying and selling a small fraction of that vignette person’s Social Security benefit stream. Between respondents, we therefore have two main treatments: (a) the information about the vignette person, randomized between “No added complexity,” “Complexity: Wide age range,” and “Complexity: Added information,” and (b) whether narrow choice bracketing is discouraged, where we randomize between “No consequence message” and “Consequence message.” In addition, we have the following six secondary randomizations. We perform two randomizations to test for anchoring, which is another indication of a lack of rational decision making (c) the starting value for the lump-sum amount ($LS = $10,000, $20,000, $30,000) and (d) the order of the two annuity valuation questions. Finally, we randomize (e) the name and gender of the primary vignette person (Mr. Jones, Mrs. Jones, Mr. Smith, Mrs. Smith) and of the secondary vignette person; (f) the Social Security benefit ($SSB = $800, $1,200, $1,600 or $2,000); (g) the order of the options shown (option with lump sum always shown first versus option with lump sum always shown last); and (h) whether the consequence message first discusses the consequences of spending wealth down quickly or slowly. The latter four manipulations are intended to verify that choices in the vignette that we assumed would be innocuous indeed do not matter for our results. All randomizations occur across subjects and are mutually orthogonal. The options within each randomization have an equal probability of being selected.

E. Data on Cognition

To investigate how the ability to value annuities varies by cognitive ability, we merge the data from our survey with existing data in the UAS, including a financial literacy survey (Lusardi and Mitchell, 2014). We also include four subtests of the Woodcock-Johnson Test of Cognitive Ability, a nationally normed test, with subtests including numeracy, number series, verbal analogies, and picture vocabulary. The first two subtests measure numerical ability, and the second two measure lexical ability. We standardize the financial literacy measure and each of the four test scores: for the main analysis, we create a cognition index from these four tests and the financial literacy measure by taking their first principal component. In the robustness section, we demonstrate the robustness of our main results to alternative measures of cognition.

III. Results

A. Baseline Sample and Randomization Check

As noted in section IIA, our baseline sample consists of respondents who answer both annuity valuation questions and have nonmissing values for the cognition and demographic variables. We investigate whether the exclusion from the baseline sample due to missing data is balanced across the two key treatment conditions (see online appendix table A3), and we find that neither the complexity treatment nor the consequence message treatment affects the likelihood of a respondent failing to answer the annuity questions (p-values: 0.322 and 0.491, respectively). The fraction of observations with missing demographic data is marginally significantly higher in the complexity treatment than in the control condition, and the fraction with missing cognition data is significantly higher in the complexity treatment than in the control condition. Since both demographic and cognition data were collected prior to randomization, these findings cannot logically be a consequence of the treatment, and we conclude they were a fluke of the randomization. There are no significant differences in the fractions with missing demographics or cognition data between the consequence treatment and the control condition.

The secondary vignette person (the vignette featured in the consequence message) was female if and only if the primary vignette person was male, and vice versa. Similarly, the secondary vignette person was named Jones if and only if the primary vignette person was named Smith, and vice versa. We did this to eliminate the possibility that the consequence message affected advice on annuity choices for the primary vignette person by respondents inferring the primary vignette person’s preferences or circumstances from information provided in the consequence message. Because the consequence message used a different person, it can only have altered the advice by the respondent through the respondent thinking differently about annuitization decisions rather than the respondent knowing more about the annuitant himself or herself.

10The secondary vignette person (the vignette featured in the consequence message) was female if and only if the primary vignette person was male, and vice versa. Similarly, the secondary vignette person was named Jones if and only if the primary vignette person was named Smith, and vice versa. We did this to eliminate the possibility that the consequence message affected advice on annuity choices for the primary vignette person by respondents inferring the primary vignette person’s preferences or circumstances from information provided in the consequence message. Because the consequence message used a different person, it can only have altered the advice by the respondent through the respondent thinking differently about annuitization decisions rather than the respondent knowing more about the annuitant himself or herself.

11The online appendix provides more detail on the construction of the cognition index and the questions used.
This figure plots the cumulative distribution function (CDF) of the sell and the buy price for a real annuity that pays out $100 a month. The sell prices are plotted for the 2,009 observations in the baseline sample for which the sell question was asked first. The buy prices are plotted for the remaining 2,015 observations in the baseline sample—those for which the buy question was asked first. This avoids the influence of anchoring on a previously asked valuation question.

We also test for balance on the control variables in the baseline sample by the two main treatments (panel B, online appendix table A3). Of the four dozen tests of differences in means across treatments for individual control variables, four are significant at the 10% level and one at the 5% level. This is roughly what one would expect by chance. Jointly, the control variables do not significantly predict the complexity treatment (p-value: 0.107) or the consequence message treatment (p-value: 0.788).

**B. Annuity Valuation Distributions and Summary Statistics**

Figure 1 shows the distribution of buy valuations for the subsample in which the buy valuation is asked first and the distribution of sell valuations for the subsample in which the sell valuation is asked first. By focusing on valuations when the question is asked first, we avoid any influence of anchoring on a previously asked valuation question. The figure clearly shows that the buy valuation is lower than the sell valuation throughout the distribution. Respondents advise our hypothetical vignette individuals to *buy* an annuity that pays $100 per month for a median price of $4,750 (SE: $180) but advise them to *sell* this annuity for a median price of $16,250 (SE: $543). This represents a statistically significant difference (two-sample Wilcoxon-Mann-Whitney rank-sum test z-statistic = 25.8, p-value < 0.001). The actuarially fair value of this annuity is roughly $15,000 at a 3% real discount rate.

Rational individuals should place a similar value on a marginal increase and a marginal decrease in the Social Security annuity. To examine the extent to which this holds in our data where we ask about a $100 change in the Social Security annuity, we calculate for each respondent the difference between the log sell price and the log buy price. Figure 2 shows the distribution of this log difference for our baseline sample and highlights two facts. First, there are large differences between buy and sell values at the individual level. Only about 10% of respondents have a buy value that is equal to their sell value, and only 40% have a buy and sell value that are within one log unit (i.e., within a factor of 2.72) of each other. In short, deviations from the predictions of the rational model for buy and sell valuations of marginal changes in Social Security benefits are substantial. Second, the distribution is not symmetric around 0: 63% have sell valuations that are higher than the buy valuation.

12 Online appendix figure A1 shows the distributions of the buy and sell valuations in the entire baseline sample which, unlike figure 1, includes responses to valuation questions that followed an earlier valuation question. The distributions are similar to those in figure 1.

13 The average of the median buy valuation and the median sell valuation is lower than the actuarially fair value. Why this is the case is not the focus of this paper’s investigation.
strictly exceed their buy valuations, whereas buy valuations strictly exceed sell valuations for about 27% of respondents. As Brown et al. (2017) explain, people may worry that they might be taken advantage of when they trade a good that they cannot value accurately. Accordingly, it can be a useful heuristic to be reluctant to trade such goods and only to sell them at a very high price (or buy them at very low price). Such a heuristic predicts that sell prices exceed buy prices whenever it is difficult to accurately determine the value of a good, as is the case with an annuity.

Status quo bias (or an endowment effect) in the level of Social Security benefits cannot explain why sell prices generally exceed buy prices. We elicit the sell price as the price for which people would be willing to sell $100 of Social Security benefits that would be received on top of the expected benefits. Someone with status quo bias would put a low price on this $100 of benefits because this amount is in addition to the status quo level of benefits. Conversely, we elicit the buy price as the price for which people would be willing to buy $100 of Social Security benefits that would bring the total benefit level back to the expected level. Thus, someone with status quo bias would place a high price on these benefits because they would return the benefit level to the status quo.

Any difference between the sell and buy prices is a deviation from the prediction of the rational model for marginal changes in Social Security benefits, whether the sell price differs from the buy price due to the reluctance-to-trade heuristic offered by Brown et al. (2017) or for other reasons. Accordingly, our measure of the deviation from rational decision making is the absolute value of the difference between the log buy price and the log sell price. We refer to this variable as the spread and use it as our main outcome variable. Table 2 presents summary statistics of the spread, and online appendix figure A2 plots its distribution. The spread is strictly positive for 90% of respondents, the median spread is 1.55, and the mean spread is 2.21. The table also shows the components of the spread, namely, the log buy price and the log sell price. Anchoring mainly affects the buy price, which is significantly higher when asked after the (generally higher) sell price is elicited. The spread is slightly higher when the sell question is asked first (2.27 versus 2.16), but this difference is only marginally significant ($p$-value: 0.079). Because the spread is measured as an absolute log difference, an increase in the spread of 0.11 (from 2.16 to 2.27) can be interpreted as the difference between the higher valued annuity and the lower valued annuity increasing by 12 ($=e^{0.11}-1$) percentage points.

As we show in online appendix table A4, demographic characteristics by themselves explain about 11% of the variation in the spread among individuals in the control group, that is, those who see the “no added complexity” vignette and do not receive the consequence message. The spread is
Table 2.—Descriptive Statistics on the Sell Price, Buy Price, and Spread

<table>
<thead>
<tr>
<th></th>
<th>Sell Question First</th>
<th>Buy Question First</th>
<th>Entire Baseline Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard Deviation</td>
<td>Mean</td>
</tr>
<tr>
<td>Sell value (log)</td>
<td>9.65</td>
<td>1.53</td>
<td>9.71</td>
</tr>
<tr>
<td>Buy value (log)</td>
<td>9.06</td>
<td>2.43</td>
<td>8.28</td>
</tr>
<tr>
<td>Sell-buy spread</td>
<td>2.27</td>
<td>2.04</td>
<td>2.16</td>
</tr>
<tr>
<td>N</td>
<td>2,009</td>
<td></td>
<td>2,051</td>
</tr>
</tbody>
</table>

Whether the buy valuation or sell valuation was asked first was randomized for each respondent. The p-value corresponds to the test that the mean in column 1 is equal to the mean in column 2. The sell-buy spread is defined as the absolute difference between the log sell price and the log buy price for a real annuity stream of $100 per month.

Table 3.—Treatment Effects on the Sell-Buy Spread and Its Components

<table>
<thead>
<tr>
<th>Explanatory Variables</th>
<th>(1) Sell-Buy Spread</th>
<th>(2) Sell Price (log)</th>
<th>(3) Buy Price (log)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity treatment</td>
<td>0.131**</td>
<td>0.050</td>
<td>-0.137**</td>
</tr>
<tr>
<td>Consequence message treatment</td>
<td>-0.141**</td>
<td>0.011</td>
<td>0.133</td>
</tr>
<tr>
<td>Cognition index</td>
<td>-0.788***</td>
<td>-0.188***</td>
<td>0.098***</td>
</tr>
<tr>
<td>Sell question first</td>
<td>0.166***</td>
<td>-0.043</td>
<td>0.777***</td>
</tr>
<tr>
<td>p-value on lump-sum starting values</td>
<td>0.623</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>p-value on lump-sum shown first</td>
<td>0.633</td>
<td>0.425</td>
<td>0.316</td>
</tr>
<tr>
<td>p-value on SS benefit amounts</td>
<td>0.249</td>
<td>0.563</td>
<td>0.000</td>
</tr>
<tr>
<td>p-value on vignette names</td>
<td>0.375</td>
<td>0.552</td>
<td>0.033</td>
</tr>
<tr>
<td>Demographic controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.157</td>
<td>0.035</td>
<td>0.067</td>
</tr>
<tr>
<td>N</td>
<td>4,060</td>
<td>4,060</td>
<td>4,060</td>
</tr>
</tbody>
</table>

The sell-buy spread is defined as the absolute difference between the log sell price and the log buy price for a real annuity stream of $100 per month. Each column displays the results from a single OLS regression, with the dependent variable listed in the column heading. The demographic controls consist of a quadratic age, a female dummy, a married dummy, three race/ethnicity dummies, four educational attainment dummies, four dummies for household income categories, three household-size dummies, and a dummy for education present in the household. Coefficient estimates on the secondary experimental treatments and the control variables are reported in appendix table A5. Robust standard errors are in parentheses. Significant at **10%, ***5%, ****1%.

significantly higher for women, non-Hispanic blacks, and those with lower education levels. The most powerful predictor of the spread is the cognition index; those with higher levels of cognition have significantly lower spreads. By itself, the cognition index can explain 16% of the variation in the spread in the control group. If we regress the spread on both the cognition index and the demographics, the $R^2$ rises to 19%, and the cognition index is the strongest and most significant predictor of the spread. The only demographic characteristic that is significant at the 5% level in this regression is the indicator for being Hispanic, which implies that Hispanics have a lower spread than what would be predicted based on their cognition score and their other demographic characteristics.

Our findings on the discrepancy between buy and sell valuations are in line with the results of Brown et al. (2017), who asked respondents for how much they themselves would buy or sell an annuity that paid them $100 per month. This similarity is reassuring, as it suggests that our elicitation of valuation advice to a vignette person (rather than asking about respondents’ own valuations) does not meaningfully alter the responses. A further similarity is that we also find that the log buy and the log sell valuations are negatively correlated (correlation coefficient: $-0.11$, p-value $< 0.001$). Our use of vignettes allows us to vary the complexity of the annuity by experimentally altering the dispersion of ages of death, which would not be ethically feasible when asking about an annuity tied to the respondent’s own life.

C. Treatment Effects

In table 3, we investigate our two main research questions. The first asks whether complexity inhibits respondents’ ability to value an annuity stream. The second asks whether narrow choice bracketing contributes to respondents’ difficulty in valuing the annuity. We measure respondents’ inability to value an annuity by the spread between their sell and buy valuations because the spread should be approximately 0 for fully rational respondents. In all regressions, we control for the experimental manipulations, the cognition index, and a common set of control variables (see panel b, online appendix table A3). In table 3, we report only the coefficients of interest (the full set of coefficient estimates is provided in online appendix table A5).

The estimate in the first row of column 1 shows that the complexity treatment increases the sell-buy spread by 0.131, implying a 14% ($= 0.131 \times 1$) increase in the ratio of the higher-valued to the lower-valued annuity. To our knowledge, this is the first causal evidence that the complexity of an annuity

14The negative correlation and the discrepancy between buy and sell prices are also consistent with the results of Chapman et al. (2017), who elicit buy and sell prices for a monetary lottery in an incentivized way and show that these prices are persistent within individuals over time and that the discrepancy between buy and sell prices is not due to measurement error.

15We do not control for the order in which the two blocks of consequence message treatment were shown because this variable is available for only half the sample. Within the half of the sample for which this order was randomized, the order has no significant effect on the spread (p-value: 0.758).
choice affects individuals’ reported annuity valuations. The fact that complexity increases the spread between the buy and sell price indicates that complexity reduces individuals’ ability to value an annuity accurately. The next two columns show the separate effects of the complexity treatment on the buy and sell price. While the estimates seem to indicate that the complexity treatment primarily operates on the buy price, and hence it reduces the average of the log sell and buy price, this is not a valid interpretation as we cannot reject that increase in the sell price and the decrease in the buy price are the same in absolute value (p-value: 0.302). We also evaluate whether the two types of complexity treatments (wide age range versus added information) have different effects on the spread. As reported in online appendix table A6, this is not the case (p-value: 0.646), so we therefore pool the two complexity treatments.

The second row of table 3 shows the treatment effects of the consequence message. The consequence message decreases the sell-buy spread by 0.141. This means that inducing respondents to think about how to spend down savings during retirement causes them to report an annuity sell price and a buy price that are closer together, which is consistent with being more able to value annuities rationally. Apparently, the consequence message reduces the degree to which respondents consider annuitization and the spending down of assets during retirement as two separate decisions, a form of narrow choice bracketing. The consequence message does move the buy and sell value closer by 15 (= e0.141-1) percentage points, but this still leaves a substantial spread of 2.21 − 0.14 = 2.07 log units among respondents who received the consequence message. In short, decision making among those who receive the consequence message is still far from rational, given that the spread remains well above 0. The next two columns show that the consequence message has virtually no effect on the sell price but significantly increases the buy price. In fact, it marginally significantly increases the average of the log buy and sell price (p-value 0.073), suggesting that the consequence message not only increases the rationality of the annuity valuations but also raises the levels. The latter finding is what one would expect when people jointly consider the asset decumulation decision and how to value the lifetime income stream. In particular, annuities remove uncertainty in consumption associated with asset decumulation in the face of uncertain life spans.

The third row shows that the cognition index is a very strong predictor of the sell-buy spread, such that a standard deviation increase in the cognition index narrows the sell-buy spread by 0.788. This underscores the conclusion that cognitive limitations play an important role in people’s inability to value an annuity. This limitation had been previously established in a different setting by Brown et al. (2017), but we now have causal evidence on two mechanisms by which cognition affects people’s ability to value annuities: narrow choice bracketing and the complexity of the annuity choice. The effect of cognition also allows us to put the magnitudes of the treatment effects in perspective. Each of our two treatments, which by coincidence have the same absolute magnitude of around 0.14, has the same effect on the spread as roughly a 17% (= 0.14/0.79) of a standard deviation change in cognitive ability.

The remaining rows examine the effects of our secondary randomizations. Consistent with earlier findings in the literature, and indicative of less than fully rational decision making, we find significant effects of anchoring. When we ask the sell valuation first (which typically has a higher valuation than the buy valuation), the respondent’s buy valuation is significantly higher, consistent with the buy valuation being anchored on the sell valuation. We find no significant anchoring of the sell price on the buy price when the latter is asked first. The starting values ($10,000, $20,000, or $30,000) of the lump-sum amount used in the annuity value elicitation procedure also have a strong effect on the valuation reported; in fact, we can reject at the 1% level that the starting value has no effect on the sell price or the buy price. The starting value has a similar effect on the sell and buy price, resulting in no significant net effect on the spread. The remaining randomizations cover the various choices we made in the design of the experiment (whether the lump-sum amount is the first or second choice, the monthly Social Security benefit amount, and the name of the vignette person). We anticipated that these choices would be innocuous, but the randomizations allow us to test whether outcomes indeed are insensitive to them. The last three rows show that these choices have no significant effects on our main outcome variable, the sell-buy spread. With the exceptions of the effects of the vignette name and the benefit amount on the buy price, these choices also do not affect the sell or buy price. To alleviate concerns about multiple hypotheses testing, we also test whether our two key experimental manipulations, the consequence message and the complexity treatment, are jointly 0; we reject this hypothesis with a p-value of 0.0106. The p-value becomes 0.0256 if we do not pool the complexity treatment, that is, when we test that the consequence message, the wide-age-range complexity treatment, and the extra-information complexity treatment are jointly 0. If we include all the secondary experimental manipulations in the joint test, we can reject that all treatment effects are jointly

17One might expect that people with an initially higher Social Security benefit place a lower value on a $100 change in Social Security benefits, since they are already more highly annuitized. To test this, we run an alternative specification in which the baseline Social Security benefit amount is included as a linear control instead of as a set of dummy variables. Both the buy and sell value decline in the baseline amount of Social Security benefits. The effect is not significant for the sell value (p-value 0.145), but there is a significant 2.5% decline in the buy value for each additional $100 in baseline Social Security benefits.
0 with a *p*-value of 0.0098 when the complexity treatments are pooled and with a *p*-value of 0.0148 when the complexity treatments are separated out.

What would annuity valuations be if we had an intervention sufficiently powerful to cause the mean log sell price and the mean log buy price to be equal (so there is no deviation from rationality at the mean)? We can get a rough answer to this question by extrapolating the effects of each of our two main experimental interventions. The mean log difference between the sell and buy price is 1.01 (see figure 2), and the consequence message moves log sell and buy price closer by 0.122 (= 0.133 − 0.011; see columns 2 and 3 of table 3). Thus, a treatment about 8−1.01/0.122 times more powerful than our current consequence message would close the gap between the mean log sell and buy price. At that level of treatment, the median sell and buy price would be predicted to be about $17,000. Similarly, we can extrapolate the complexity treatment in the direction of making the problem less complex, such that the sell and buy price coincide. This would require reducing complexity by about five times the amount of complexity added by our complexity treatment. The resulting sell and buy price would then be predicted to be about $12,000. These point estimates obviously rely on a substantial extrapolation, and therefore they should be taken only as suggestive. Nevertheless, it is noteworthy that a simple average of these two predicted valuations at treatments sufficiently powerful to eliminate the discrepancy between the buy and sell price is quite close to the actuarily fair value (of about $15,000).

### D. Heterogeneous Treatment Effects

In table 4, we explore whether the impact of our two main treatments varies across respondent subgroups. The first column examines heterogeneity in the effect of the complexity treatment, and the second column investigates whether the consequence message has different effects across subgroups. For each specification, we create two subgroups that are as close as possible in size to each other in order to maximize statistical power.

The first two specifications examine interaction effects between our treatments. One might expect that the complexity treatment has a greater impact on the spread when people engage in narrow choice bracketing because they do not recognize how annuities help in the asset drawdown process. In line with this prediction, the point estimate of the complexity treatment is larger for respondents who receive no consequence message than for those who do; nevertheless, this difference is not statistically significant (*p*-value: 0.408). The second specification is the flip side of the first, asking whether the consequence message has a greater impact on persons exposed to the complexity treatment. While the point estimates do go in this direction, this effect is again not significant (and the *p*-value is the same as in the first specification by construction).

The third specification shows that we do not detect significant heterogeneity in either treatment effect by level of cognition. The last specification splits the estimates by the randomly assigned level of Social Security benefits. A $100 change in Social Security benefits is closer to a marginal change for someone with monthly benefits of $2,000 than for someone with monthly benefits of $800. The stability of treatment effects by level of benefits helps alleviate concerns.

---

**Table 4.—Heterogeneity in Treatment Effects**

<table>
<thead>
<tr>
<th>Specification</th>
<th>Complexity Treatment</th>
<th></th>
<th>Consequence Message Treatment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>(S.E.)</td>
<td><em>p</em>-value</td>
<td>Coeff.</td>
</tr>
<tr>
<td>1. By Consequence Message</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No consequence message</td>
<td>0.185**</td>
<td>(0.094)</td>
<td></td>
<td>0.185**</td>
</tr>
<tr>
<td>Consequence message</td>
<td>0.078</td>
<td>(0.089)</td>
<td></td>
<td>0.078</td>
</tr>
<tr>
<td><em>p</em>-value on test of equal coefficients</td>
<td>0.0098</td>
<td></td>
<td></td>
<td>0.0098</td>
</tr>
<tr>
<td>2. By Complexity Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No complexity treatment</td>
<td>−0.071</td>
<td>(0.104)</td>
<td></td>
<td>−0.071</td>
</tr>
<tr>
<td>Complexity treatment</td>
<td>−0.178**</td>
<td>(0.077)</td>
<td></td>
<td>−0.178**</td>
</tr>
<tr>
<td><em>p</em>-value on test of equal coefficients</td>
<td>0.0148</td>
<td></td>
<td></td>
<td>0.0148</td>
</tr>
<tr>
<td>3. By Cognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below median cognition index</td>
<td>0.132</td>
<td>(0.103)</td>
<td></td>
<td>−0.167*</td>
</tr>
<tr>
<td>Above median cognition index</td>
<td>0.133*</td>
<td>(0.077)</td>
<td></td>
<td>−0.117</td>
</tr>
<tr>
<td><em>p</em>-value on test of equal coefficients</td>
<td>0.082</td>
<td></td>
<td></td>
<td>0.082</td>
</tr>
<tr>
<td>4. By Level of Social Security Benefits</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below median ($800 or $1200)</td>
<td>0.123</td>
<td>(0.092)</td>
<td></td>
<td>−0.142</td>
</tr>
<tr>
<td>Above median ($1200 or $1600)</td>
<td>0.139</td>
<td>(0.091)</td>
<td></td>
<td>−0.140</td>
</tr>
<tr>
<td><em>p</em>-value on test of equal coefficients</td>
<td>0.0093</td>
<td></td>
<td></td>
<td>0.0093</td>
</tr>
</tbody>
</table>

*The sell-buy spread is defined as the absolute difference between the log sell price and the log buy price for a real annuity stream of $100 per month. Each row reports the results from a single OLS regression in which the two main experimental treatments are interacted with the characteristics listed in the row header. Robust standard errors are in parentheses. Significant at *10%, **5%, and ***1%.*

---

18 Online appendix table A7 examines heterogeneity by gender, education, age, and income. In none of these specifications do we find a difference in the treatment effect by demographic characteristic significant at the 5% level or better, but we recognize that we have limited statistical power to detect even reasonably large interaction effects.
that the estimates are affected by the fact that the $100 change is not literally a marginal change. Another way to address this concern is to not count small spreads as deviations from rational behavior, which could arise when a $100 change is insufficiently marginal. In online appendix table A8, we do this by setting any spreads less than 0.50 log units equal to, and we find that the estimated treatment effects are essentially unaffected.

E. Robustness

Online appendix table A8 examines the robustness of the two primary treatments to different measures of cognition, different ways of selecting the sample, different sets of controls, and transformations of the outcome variable. We find that the results on the complexity treatment are reasonably stable in magnitude but somewhat sensitive in terms of statistical significance, which falls to marginal in seven of the eighteen specification checks and disappears in two of them. This sensitivity can be traced largely to the fact that the cognition control, a very strong predictor of the spread, was not balanced across the complexity treatment and control conditions. Hence, having good controls for cognition is important for the results of the complexity treatment. By contrast, the consequence message treatment is extremely robust and remains significant at the 5% level everywhere, except for one specification where it is significant at the 10% level.

IV. Conclusion

Annuities allow people to smooth consumption in retirement when facing an uncertain age of death, yet annuity holdings are relatively low and only about 3% of individuals maximize their annual Social Security annuity payouts by delaying claiming benefits until age 70 (Social Security Administration, 2017). Although these decisions may be rational for some people, this paper investigates whether behavioral factors impede people’s annuitization choices. We do so in the context of a hypothetical choice experiment on a broadly representative sample of about 4,000 adults in the United States. Such a setting confers two important advantages for our purposes. First, we can measure deviations from rational decision making by observing for each respondent both his willingness to pay to forgo a small decrease in annuitization and his willingness to accept to forgo a small increase in annuitization. Second, we can experimentally vary the complexity of the annuitization decision. We also experimentally vary whether respondents are encouraged to jointly consider the annuitization decision and the asset decumulation decision during retirement (thus discouraging narrow choice bracketing), though this treatment could in principle also be applied in nonhypothetical choice settings.

Our first main finding is that increasing the complexity of the annuity decision reduces people’s ability to value the annuity. This decreased ability manifests itself as an increase in the divergence of people’s sell and buy prices for a marginal change in annuitization. When the annuity decision becomes more complex, people tend to become more reluctant to buy or sell annuities, meaning they need greater inducements (lower buy or higher sell prices) to do so. Brown et al. (2017) document that a reluctance to trade annuities, as measured by the sell-buy price spread, is strongly negatively associated with cognitive ability, but of course, cognitive ability is not randomly assigned. In our setting, we experimentally vary the complexity of the annuitization decision to obtain the first causal evidence that more complex annuitization decisions reduce people’s ability to place a value on an annuity, as measured by the sell-buy spread. Hence, the observed low level of annuity holdings can be traced at least in part to the cognitive challenges of the complex task of valuing an annuity.

The second finding is that inducing people to think jointly about annuitization and how to draw down assets during retirement increases their ability to place a value on an annuity. We experimentally induce respondents to think about these decisions jointly by exposing them to a consequence message that explains the result of spending down assets more slowly or more rapidly during retirement. Respondents who think about this asset decumulation decision have a smaller sell-buy spread for annuities than do respondents not exposed to the consequence message. This finding suggests that narrow choice bracketing, which the consequence message counters, is one behavioral mechanism impeding people from placing a rational value on annuities.

Our results on the roles of complexity and cognitive ability offer relatively little scope for interventions to improve the quality of people’s annuitization decisions. Cognitive ability for any given person is relatively immutable, as is the complexity of the annuitization decision. While this complexity can be somewhat diminished by presenting the annuity information more transparently, most of the complexity stems from having to consider how the annuity would alter consumption streams in different states of the world, an inherently complex task. In contrast, our finding on the role of narrow choice bracketing does offer scope for interventions to improve people’s decision making about annuities. In particular, people provide more rational annuity valuations if they first consider the question of how to spend down nonannuitized wealth during retirement. We therefore conclude that annuitization decisions can be improved by inducing people to jointly consider annuitization and spending down nonannuitized wealth.

Although our experimental setting is that of a hypothetical person facing an annuity decision, we believe our results inform understanding of an important set of nearly universal decisions. It is comforting that the distribution of buy-sell spreads found using these vignettes is similar to the distribution found in earlier research in which individuals were making hypothetical decisions for themselves (Brown et al., 2017). By using Social Security as our context, we are confident that the lifelong income feature of Social Security is widely understood. We therefore believe that these results
Behavioral Impediments to Valuing Annuities

potentially generalize to any situation in which an individual must place an implicit value on a stream of annuity income, including whether to claim Social Security benefits immediately on retirement or to delay claiming, whether to accept a lump-sum payment in lieu of an annuity from an employer’s defined benefit pension plan, or whether to annuitize assets in a defined contribution plan. We think it is plausible that the behavioral mechanisms that drive the results in our setting would also operate in other markets, such as those for stocks, options, and insurance, but whether this is indeed the case should be based on research in those markets rather than on the extrapolation of our results.

Our results do not explain why average valuations are below actuarially fair levels, and thus our results should not be interpreted as fully explaining the annuity puzzle. Indeed, we do not believe that complexity and narrow choice bracketing are the only reasons that individuals are reluctant to annuitize. Nevertheless, our paper adds to the evidence that behavioral factors influence annuitization decisions, and it also provides causal evidence on two specific mechanisms: narrow choice bracketing and cognitive limitations to dealing with complex decisions. Naturally, our evidence on these two behavioral impediments to valuing annuities does not preclude other mechanisms (Brown 2009). One avenue for future investigation will be to quantify the welfare effects of behavioral deviations from rational decision making in the context of annuitization decisions.

REFERENCES


